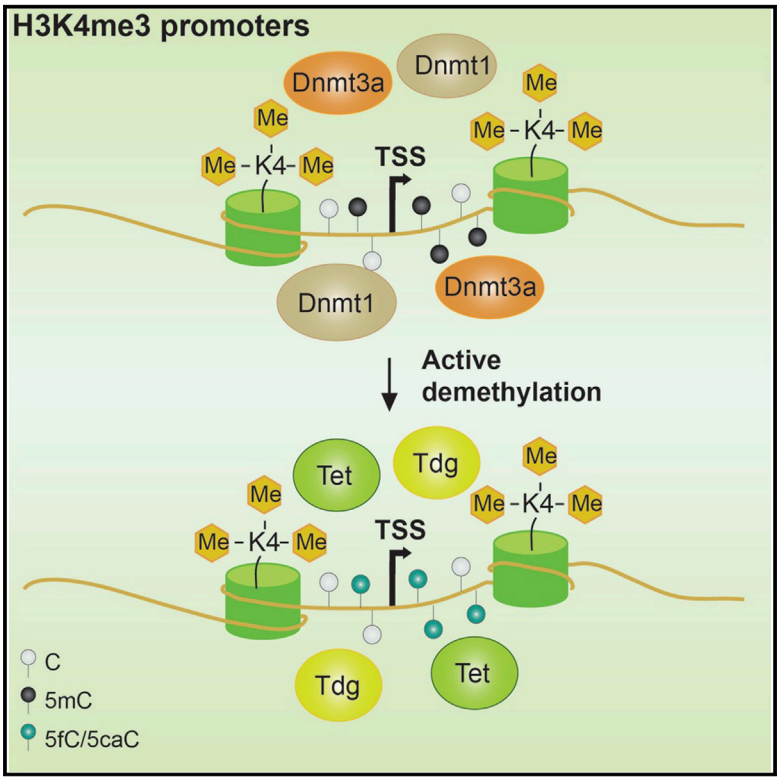
Cell Reports

Report

Single-Base Resolution Analysis of 5-Formyl and 5-Carboxyl Cytosine Reveals Promoter DNA Methylation Dynamics

Graphical Abstract



Authors

Francesco Neri, Danny Incarnato, ..., Federica Dal Bello, Salvatore Oliviero

Correspondence

salvatore.oliviero@hugef-torino.org

In Brief

Ten eleven translocation enzymes convert 5-methylcytosine to 5hmC, 5fC, and 5caC, promoting DNA demethylation through base excision by the thymine-DNA glycosylase. Neri et al. map 5fC and 5caC genome wide and find them on promoters of highly expressed genes, indicating dynamic methylation and demethylation.

Highlights

- We present MAB-seq, a method for mapping genome-wide 5fC and 5caC
- 5fC and 5caC occur on active promoters and enhancers
- 5fC and 5caC correlate with Tet1 and TDG, and TDG knockdown increases their levels
- Highly expressed genes are continuously involved into methylation and demethylation

Accession Numbers

GSE55660

Neri et al., 2015, Cell Reports *10*, 674–683
February 10, 2015
http://dx.doi.org/10.1016/j.celrep.2015.01.008

CellPress

OPEN ACCESS
CellPress

Cell Reports
Report

Single-Base Resolution Analysis of 5-Formyl and 5-Carboxyl Cytosine Reveals Promoter DNA Methylation Dynamics

Francesco Neri,[1,5] Danny Incarnato,[1,2,5] Anna Krepelova,[1] Stefania Rapelli,[1,2] Francesca Anselmi,[1] Caterina Parlato,[1] Claudio Medana,[3] Federica Dal Bello,[3] and Salvatore Oliviero[1,4,*]
[1]Human Genetics Foundation (HuGeF), via Nizza 52, 10126 Torino, Italy
[2]Dipartimento di Biotecnologie Chimica e Farmacia, Università di Siena, via Fiorentina 1, 53100 Siena, Italy
[3]Dipartimento di Biotecnologie Molecolari e Scienze per la Salute, Università di Torino, Via Pietro Giuria 5, 10125 Torino, Italy
[4]Dipartimento di Scienze della Vita e Biologia dei Sistemi, Università di Torino, Via Accademia Albertina 13, 10123 Torino, Italy
[5]Co-first author
*Correspondence: salvatore.oliviero@hugef-torino.org
http://dx.doi.org/10.1016/j.celrep.2015.01.008

SUMMARY

Ten eleven translocation (Tet) proteins oxidize 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC). 5fC and 5caC can be further excised by thymine-DNA glycosylase (Tdg). Here, we present a genome-wide approach, named methylation-assisted bisulfite sequencing (MAB-seq), that enables single-base resolution mapping of 5fC and 5caC and measures their abundance. Application of this method to mouse embryonic stem cells (ESCs) shows the occurrence of 5fC and 5caC residues on the hypomethylated promoters of highly expressed genes, which is increased upon Tdg silencing, revealing active DNA demethylation on these promoters. Genome-wide mapping of Tdg reveals extensive colocalization with Tet1 on active promoters. These regions were found to be methylated by Dnmt1 and Dnmt3a and demethylated by a Tet-dependent mechanism. Our work demonstrates the DNA methylation dynamics that occurs on the promoters of the expressed genes and provides a genomic reference map of 5fC and 5caC in ESCs.

INTRODUCTION

DNA methylation on position 5 of cytosine (5mC) is the most conserved epigenetic modification from plants to animals (Feng et al., 2010) and is catalyzed by DNA methyltransferases (Dnmts). 5mC is necessary for the correct embryonic development (Bestor, 2000; Fouse et al., 2008; Li, 2002; Neri et al., 2013b; Reik and Walter, 2001), and its deregulation has been linked to cancer development and progression (Costello et al., 2000; Easwaran et al., 2012; Jones, 2002; Neri et al., 2014; Teng et al., 2011). On the other hand, active DNA demethylation is mediated by ten eleven translocation (Tet) proteins, which progressively oxidize 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC), and by the thymine-DNA glycosidase (Tdg) protein that binds and excises 5fC and 5caC residues to allow the restoration of unmodified C by the base excision and repair machinery (Cortázar et al., 2011; Cortellino et al., 2011; He et al., 2011; Ito et al., 2010, 2011; Maiti and Drohat, 2011; Neri et al., 2013a; Williams et al., 2011; Wu et al., 2011). Several methods, such as bisulfite sequencing (bisulfite-seq), Tet-assisted bisulfite sequencing (TAB-seq), and oxidative bisulfite sequencing, have been established to map 5mC and 5hmC residues at single-base resolution on a genome-wide scale (Booth et al., 2012; Meissner et al., 2008; Yu et al., 2012). Other methods have been developed to map 5fC and 5caC, which revealed the presence of these two DNA modifications on enhancers, repetitive elements, and exons (Raiber et al., 2012; Shen et al., 2013; Song et al., 2013). These later methods, which are based on affinity purification approaches, lack the resolution power to map single cytosines and to quantify the relative abundance of the modification. Single-base resolution analysis of 5fC (5fc chemical-assisted bisulfite sequencing and reduced bisulfite sequencing) has been performed using chemical methods, although only on reduced genomes (Booth et al., 2014; Song et al., 2013). Moreover, it has been developed recently a method for single-base resolution of 5caC, even though its suitability for genome-wide experiments has not been investigated (Lu et al., 2013). We here introduce a high-throughput method, named methylation-assisted bisulfite sequencing (MAB-seq), that enables genome-wide mapping at single-base resolution and quantitation of the 5fC and 5caC residues. Application of this method to mouse embryonic stem cells (ESCs) revealed the occurrence of 5fC and 5caC residues on the hypomethylated promoters of highly expressed genes, revealing an active DNA demethylation dynamics occurring on these promoters.

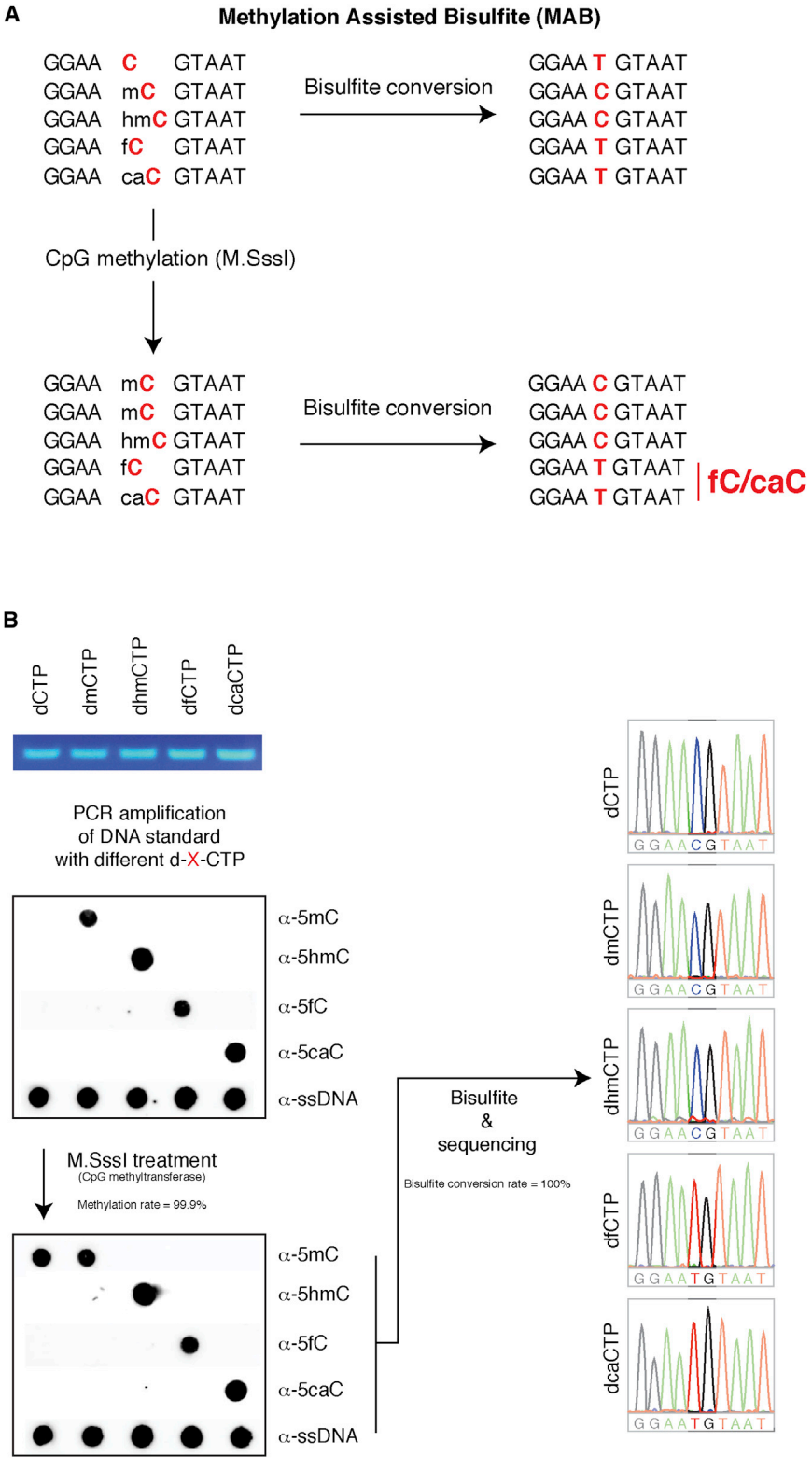


Figure 1. MAB Method Enables Highly Efficient Detection of 5fC and 5caC

(A) Bisulfite treatment of DNA leads to deamination of cytosine, 5fC, and 5caC, whereas the 5-methyl and 5-hydroxymethyl cytosine are protected. Treatment of DNA with the S-Adenosyl-methionine-dependent CpG methyltransferase M.SssI prior to bisulfite treatment leads to the conversion of unmodified cytosines to 5-methyl cytosine, thus enabling the specific detection of 5fC and 5caC residues.

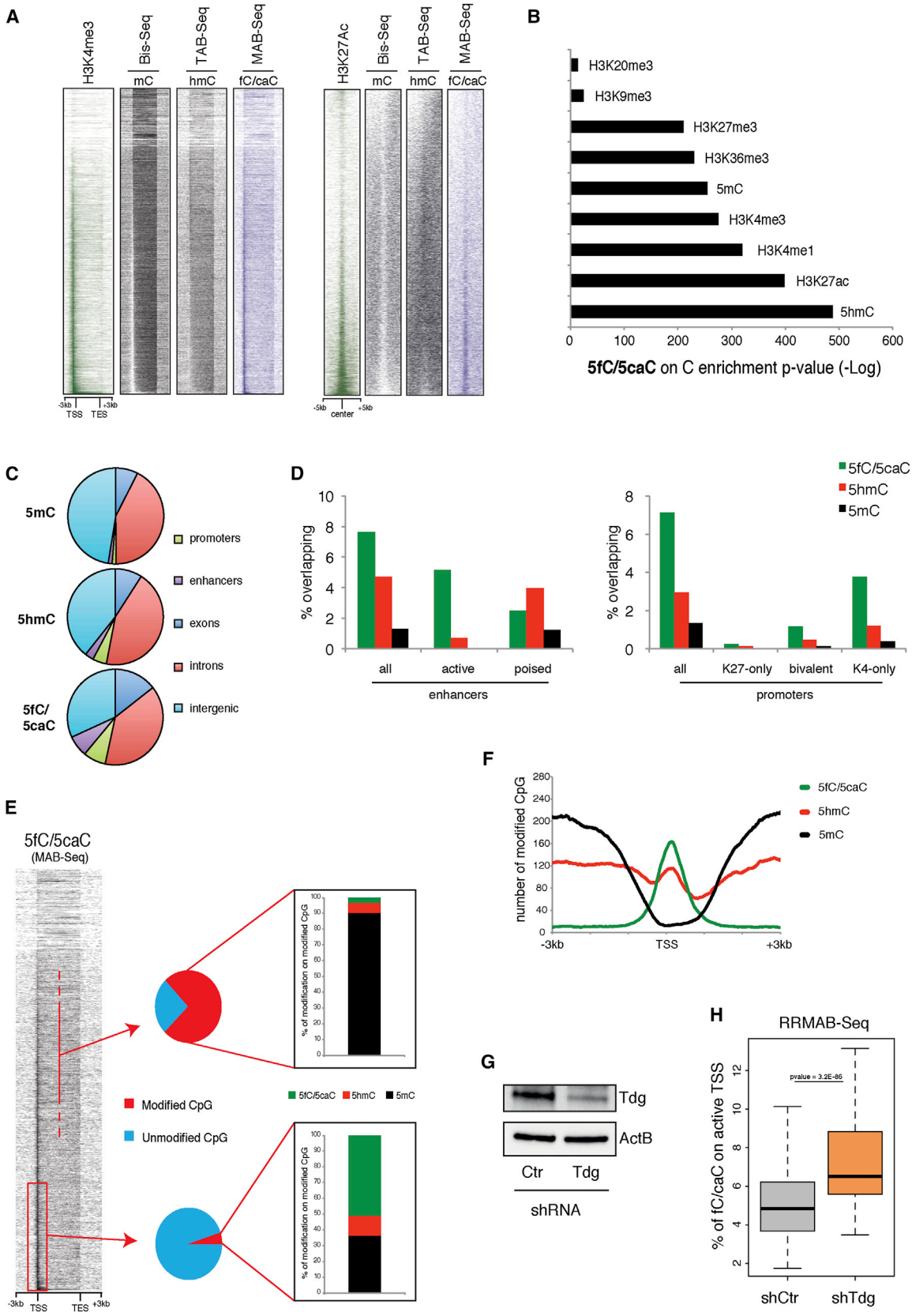
(B) Validation of MAB on five DNA standards containing, respectively, only unmodified, 5-methyl, 5-hydroxymethyl, 5-formyl, or 5-carboxyl cytosine residues. Dot blot analysis reveals complete conversion of unmodified cytosines to 5-methyl cytosines following M.SssI treatment. After bisulfite treatment, Sanger sequencing detects 5fC and 5caC as T, whereas unmodified C (methylated by M.SssI), 5mC, and 5hmC are detected as C.

RESULTS

MAB-Seq Allows Genome-wide Mapping of 5fC and 5caC Residues

We have developed a technique, named MAB-seq, that enables the identification and quantitation of the relative abundance of 5fC and 5caC residues on a genome-wide scale. Our method is based on the reactivity of these two oxidized cytosine residues to bisulfite treatment (Yu et al., 2012). We employed the CpG methyltransferase M.SssI enzyme to specifically methylate the unmodified cytosine residues, thus protecting them against the conversion by bisulfite treatment (Figure 1A). Because 5fC and 5caC residues are readily converted to uracil (U) by bisulfite like regular unmodified cytosines, the prior M.SssI treatment results in the conversion only of 5fC and 5caC residues. To calculate the efficiency of these reactions, we generated by PCR five standards, containing the different modified cytosine residues (5mC, 5hmC, 5fC, and 5caC) or unmodified cytosines, and subjected them to M.SssI and bisulfite treatment, followed by traditional Sanger sequencing. The bisulfite conversion rate was estimated from the number of 5fC and 5caC residues that underwent conversion to U and the M.SssI methylation rate from the number of unmodified Cs that were unaffected by bisulfite conversion. The extremely high efficiencies observed for both treatments (>99.9%; Figure 1B) suggested the suitability of this approach for a genome-scale analysis.

Moreover, by loss-of-function experiments, we identify Dnmt1, Dnmt3a, and Tet1/Tet2 proteins as the key players of this mechanism.



(legend on next page)

To this end, we applied MAB-seq to genomic DNA from E14 mouse embryonic stem cells and produced more than 1×10^9 sequencing reads, which is suitable for a whole-genome analysis (Table S1). To estimate the M.SssI methylation efficiency for our experiment, genomic DNA from ESCs was spiked in with *Escherichia coli* genomic DNA, which has been previously shown to lack CpG methylation and to be appropriate for bisulfite sequencing analysis (Gu et al., 2011; Kahramanoglou et al., 2012). To confidently detect 5fC and 5caC-modified bases, we used a binomial distribution, as previously described (Yu et al., 2012), with N as the sequencing depth at the converted C site and p as the cytosine nonmethylation rate. We considered only the called 5fC and 5caC sites with sequencing depth >50×, and p value < 0.001 (see Experimental Procedures). This resulted in the identification of about 3.3×10^5 5fC and 5caC sites, with a percentage of modification distributed around 8%–10% (Figure S1A).

Collectively, our analysis provided a genome-wide map of 5fC and 5caC in the ESCs genome at single-base resolution.

5fC and 5caC Are Enriched on Active Transcriptional Start Sites and Enhancers in ESCs

We next sought to determine how the 5fC and 5caC sites are distributed across the ESCs genome. Intersection of 5fC and 5caC sites coordinates with those of different genomic features revealed their enrichment on enhancers, exons, and repetitive regions (such as satellites), in agreement with three recently published studies based on affinity purification methods (Raiber et al., 2012; Shen et al., 2013; Song et al., 2013; Figures S1B–S1D). Moreover, 5fC and 5caC presented no, or low, symmetry with respect to the opposite strand, showing a behavior similar to 5hmC (Yu et al., 2012; Figures S1E and S1F).

Interestingly, MAB-seq identified a high enrichment of 5fC and 5caC sites on active promoters, especially on the transcriptional start sites of H3K4me3-only genes (Figures 2A–2C and S1B). Comparison of 5fC and 5caC with 5mC and 5hmC revealed that gene transcriptional start sites (TSSs) are more enriched for 5fC and 5caC modification with respect to 5mC and 5hmC. This correlation increased with the level of gene promoter accessibility, marked by the progressive gain of H3K4me3 mark and loss of H3K27me3 mark (Figures 2D and S2A).

Comparison between our analysis and affinity-purification-based methods revealed that only our technique is able to detect the 5fC and 5caC on the TSSs of active genes (Figures S2C and S2D). The presence of these modifications on H3K4me3-positive TSSs suggests an active DNA methylation/demethylation dynamics occurring on the promoters of actively transcribed genes in ESCs. An analogous distribution pattern was also observed on enhancers. We observed a progressive increase of 5fC and 5caC levels in correlation with the progressive enrichment of the H3K27Ac histone mark (Figures 2A–2D and S2B).

5fC and 5caC Are Bound and Excised by Tdg

We next ordered the genes by their expression level, as calculated from RNA sequencing analysis, and measured the relative abundances of each cytosine derivative on the promoter (±1 kb around the TSS) of the genes in the upper quartile of expression, compared to gene bodies. This analysis showed that modified cytosine residues around the TSS of actively transcribed genes are shifted toward a higher oxidative state, of which approximately the 50% was represented by 5fC and 5caC (Figures 2E and 2F).

To understand whether this 5mC oxidation could be finalized to its excision, and thus to the demethylation of the cytosine residues, we performed reduced representation MAB-seq (RRMAB-seq) in control and Tdg-silenced ESCs (Figure 2G). To confirm Tdg knockdown, we evaluated the reduction of its enzymatic activity and relative increase of 5fC level by colorimetric ELISA assays (Figures S3A and S3B). RRMAB-seq is based on the pretreatment of genomic DNA with the enzyme MspI, which recognizes and cuts the CCGG consensus, thus allowing to enrich CpG-rich regions, like gene promoters. Tdg knockdown resulted in a significant increase of 5fC and 5caC levels around the TSS (±1 kb) of highly expressed genes, demonstrating that an active Tdg-dependent 5fC and 5caC excision mechanism takes place on these regions (Figure 2H).

To further confirm this mechanism, we mapped the localization of Tdg on the ESCs genome. To this end, we performed chromatin immunoprecipitation, followed by deep sequencing (ChIP-seq) of Tdg, using a streptavidin-biotin affinity approach, in an ESC line expressing a biotin-tagged Tdg (Figure 3A). ChIP-seq analysis unveiled a strong enrichment of Tdg on the TSSs, with a striking preference for the genes enriched for H3K4me3, showing a distribution analogous to that of 5fC and 5caC and Tet1 (Figures 3B, S3C, and S3D). Pearson correlation analysis on a subset of ESCs-specific transcriptional factors, and histone modifications, revealed that Tdg is associated with epigenetic features of active TSSs (H3K4me3 and H3K9ac) and active enhancers (H3K4me1 and H3K27ac; Figures 3C, S3E, and S3F). Moreover, a stronger correlation for Tdg was

Figure 2. 5fC and 5caC Are Enriched on Active Promoters and Enhancers

(A) Heatmaps of 5fC and 5caC on genes (left) and enhancers (right) compared to that of 5mC and 5hmC. Genes and enhancers are ascending sorted according to levels of H3K4me3 and H3K27ac, respectively.

(B) Enrichment of 5fC and 5caC versus unmodified C on different epigenetic features. p values are given by chi-square statistics.

(C) Pie charts showing the genome-wide distribution of 5mC, 5hmC, and 5fC/5caC. Proceeding from a lower to a highly oxidized C state (5mC → 5fC/5caC), modified C residues become more enriched inside promoters, enhancers, and exonic regions.

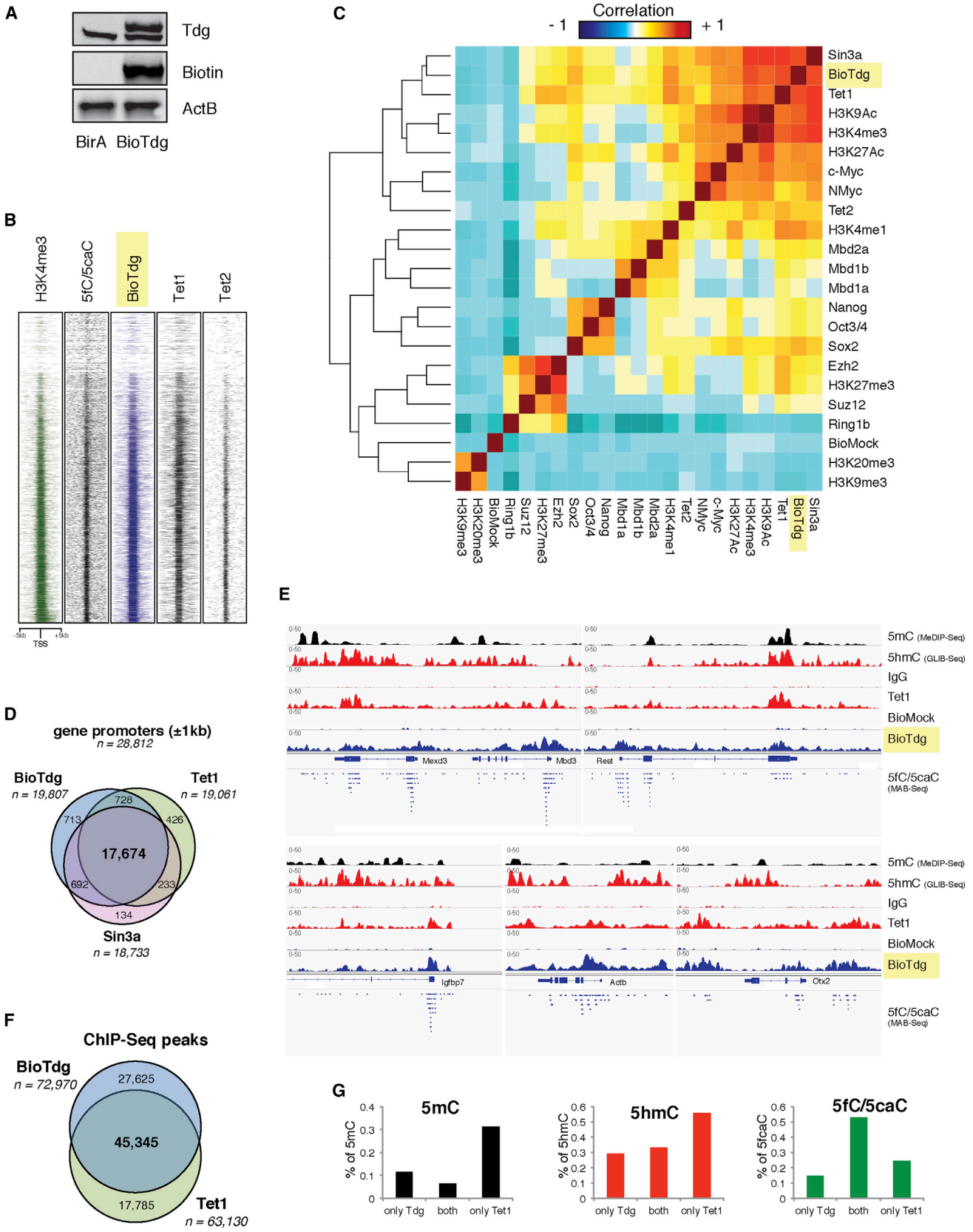
(D) Genome-wide distribution of 5mC, 5hmC, and 5fC/5caC within different enhancer (left) and promoter (right) classes.

(E) Relative fraction of C derivatives within gene bodies (top) and highly expressed genes (upper quartile), with respect to unmodified C residues. Genes are sorted ascending by expression level.

(F) Mean distribution of 5mC, 5hmC, and 5fC/5caC on a window of ±3 kb around the TSS of highly expressed genes (upper quartile).

(G) Western blot of ESC transfected either with a control small hairpin RNA (shRNA) (shCtr) or a shRNA targeting Tdg. β-actin is used as loading control.

(H) Box plot of the percentage of 5fC and 5caC on highly expressed genes (upper quartile) in Tdg knockdown versus control, as revealed by reduced representation MAB-seq (RRMAB-seq). p value is given by Wilcoxon rank sum test statistics.



(legend on next page)

observed with Tet1 and Sin3a, as demonstrated by the colocalization of these three factors on more than 90% of bound genes (Figure 3D).

Genomic views of ChIP-seq tags showed similar profiles between Tdg and Tet1, with some differences in their relative abundances. On the TSSs, Tdg and 5fC and 5caC were enriched concomitantly with a depletion of 5mC and 5hmC. In contrast, on gene bodies, especially on exons and transcriptional end sites, Tet1 showed a more-abundant binding, concordantly with a higher enrichment of 5hmC and sometimes of 5mC (Figure 3E). Tdg and Tet1 shared more than 70% of ChIP-seq peaks, indicating a significant colocalization on the ESC genome (Figure 3F). Collectively, these data suggest that Tet1 and Tdg relative abundance controls the dynamics of 5mC oxidation and excision. Accordingly, the regions targeted by both enzymes were enriched for 5fC and 5caC and were hypomethylated; whereas regions preferentially bound by Tet1 showed an enhancement of 5mC and 5hmC levels (Figure 3G).

Taken together, these results suggest that the promoters of actively transcribed genes, which are believed to be unmethylated, are in fact targets of DNA methylation.

Active TSSs Are Methylated by Dnmt1/Dnmt3a and Demethylated by Tet1/Tet2

To verify that active TSSs are actually targeted by DNA methyltransferases, we stratified all cytosines by their level of methylation, using a high-coverage data set of bisulfite-seq in ESCs (Habibi et al., 2013), and subtracted the relative value of 5hmC obtained by TAB-seq analysis (Yu et al., 2012). Cytosines were split into ten subsets, according to their modification level (0%–100%; offset: 10%), and plotted on the TSS (±5 kb) of all genes, sorted by their H3K4me3 level (Figure 4A). Notably, cytosines with a lower percentage of methylation (0%–20%) were enriched on the TSS, with a direct correlation to the level of gene expression. This TSS-centric distribution of 5mC is lost upon increasing the cytosines methylation percentage and progressively shifts toward more distal regions. Thus, highly methylated cytosines are absent on the TSS of expressed genes, but the TSS-centric distribution observed at lower methylation levels suggests that DNA methylation events actually occur on these regions. This 5mC distribution pattern also marks the enhancers and can also be observed for 5hmC (Figures S4A–S4C).

To demonstrate that the presence of 5fC and 5caC and Tdg on actively transcribed genes is actually aimed to an active demethylation, we measured the level of DNA methylation on these regions in control and Tet1/Tet2 double-knockdown cells by reduced representation bisulfite sequencing (RRBS) analysis that allows a high coverage on regions marked by CpG islands, such as promoters of active genes (Figures 4B, 4C, S4D, and S4E). Loss of Tet1/Tet2 proteins resulted in a significant increase of 5mC levels, demonstrating a functional enzymatic activity of Tet1/Tet2 on the promoters of actively transcribed genes (Figure 4D).

To further identify the methyltransferases responsible for the methylation of these loci, we silenced Dnmts in ESCs (Figure 4E). Analysis of promoter methylation by RRBS following Dnmts knockdown showed a significant reduction of DNA methylation levels in Dnmt1 and Dnmt3a knockdown cells, whereas loss of Dnmt3b did not affect the 5mC levels on the promoter of highly expressed genes (Figures 4F, S4F, and S4G). Collectively, these results demonstrate that Dnmt1 and Dnmt3a, but not Dnmt3b, are responsible for the methylation on these genomic regions.

DISCUSSION

We here provided a method (MAB-seq) for the mapping and quantitation of 5fC and 5caC in the mouse genome. This method shows an increased sensitivity with respect to affinity-based purification techniques and the ability to map directly the 5fC and 5caC modifications with respect to single-base resolution subtractive approaches. Despite the low abundance of 5fC and 5caC, the direct readout provided by MAB-seq identifies these modifications without the need for multiple high-coverage sequencing experiments. Because 5fC and 5caC oxidative states are catalyzed by the same enzyme and are both removed by Tdg, the profiling in a single sequencing analysis of these modifications enables a fast and streamlined mapping of the genomic sites that undergo active DNA demethylation.

By applying MAB-seq to ESCs, we here provided a whole-genome map of 5fC and 5caC at single-base resolution. Besides the detection of 5fC and 5caC at enhancers and repetitive regions, in agreement with affinity purification methods, we observed a significant enrichment of 5fC and 5caC on the promoters of actively transcribed genes due to the greater sensitivity of the MAB-seq technique with respect to affinity-based methods. The presence of 5fC and 5caC on H3K4me3-positive regions reveals an active demethylation process on active promoters. Stratification of CpGs by their methylation level (Figure 4A) revealed the presence of 5mCs around the majority of the active TSSs, suggesting the occurrence of active methylation/demethylation mechanisms on H3K4me3 regions, explaining the presence of Tet1/Tet2 proteins on these regions. Moreover, by loss-of-function experiments, we identified Dnmt1 and Dnmt3a, but not Dnmt3b, as responsible for the methylation at active TSSs. The involvement of Dnmt1 in this process suggests either that methylation on these regions is more stable than

Figure 3. Tdg Colocalizes with 5fC and 5caC and Tet1 Genome Wide

(A) Western blot of biotin-tagged Tdg in BirA-expressing ESCs. β-actin is used as loading control.

(B) Heatmaps of Tdg-bound regions on a window of ±5 kb around the TSS of genes. We also report for comparison heatmaps of 5fC and 5caC (this study), Tet1, and Tet2. Genes are ascending sorted according to levels of H3K4me3.

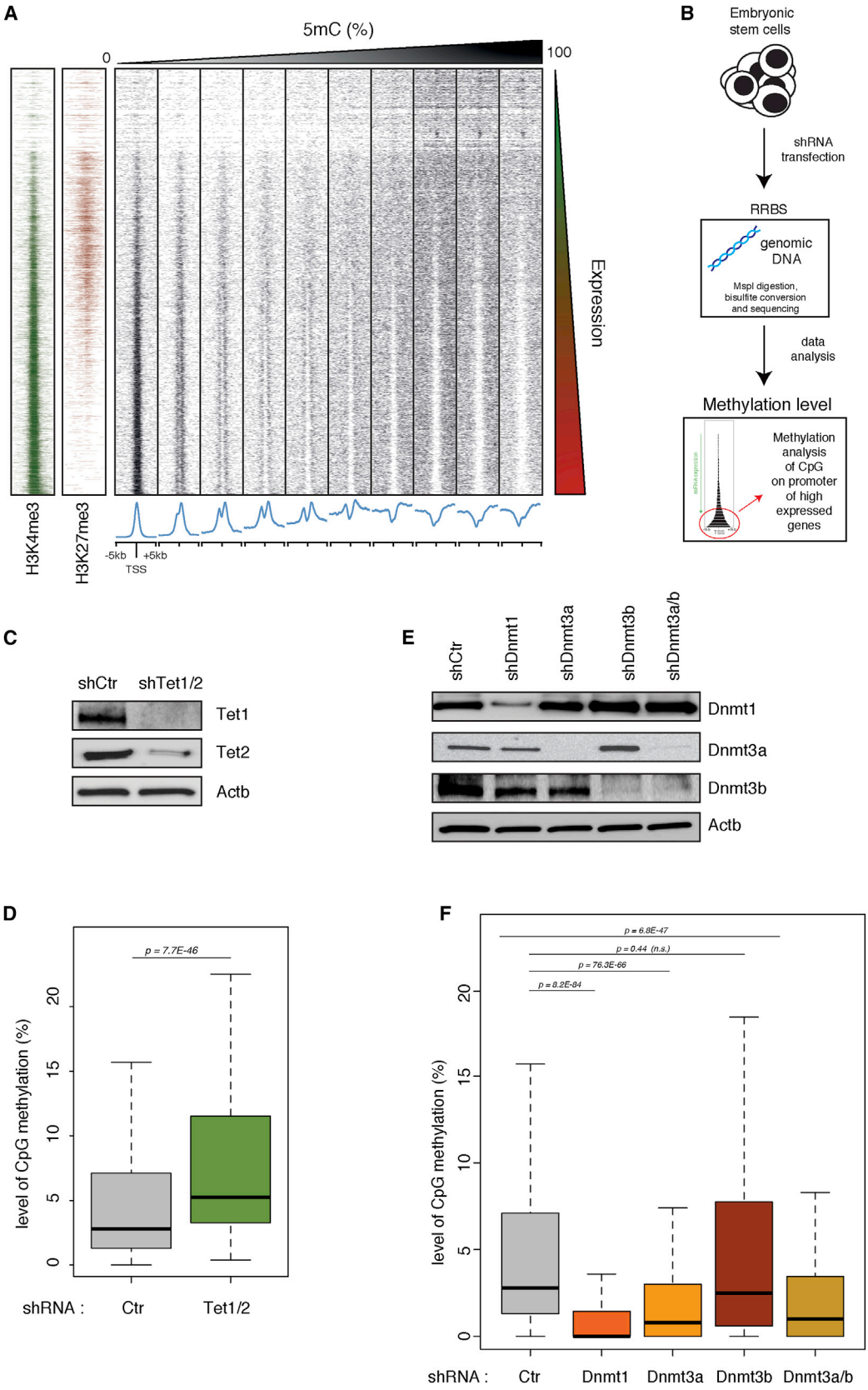
(C) Hierarchical clustering of pairwise Pearson correlation on Bio-Tdg (and Bio-Mock control) and third-party ChIP-seq data sets in ESCs.

(D) Venn diagram illustrating the overlap between gene promoters targeted by Tdg, Tet1, and Sin3a.

(E) Example genomic views of Tdg targets in ESCs. IgG, immunoglobulin G.

(F) Venn diagram illustrating the overlap between Tdg and Tet1 ChIP-seq peaks.

(G) Percentages of 5mC, 5hmC, and 5fC and 5caC on gene promoters targeted exclusively by Tet1 or by Tdg and targeted by both factors.



(legend on next page)

expected or, more likely, that Dnmt1 can de novo methylate these regions either alone or together with Dnmt3a (Fatemi et al., 2002; Gowher et al., 2005; Jair et al., 2006). Accordingly, the loss of Tet enzymes increased the level of DNA methylation on active TSSs. Our results demonstrate genome wide the occurrence of DNA methylation and demethylation previously shown on the TFF1 and ERα promoters (Kangaspeska et al., 2008; Métivier et al., 2008).

We also mapped by ChIP-seq the genomic localization of Tdg. This analysis showed Tdg binding on the TSSs of actively transcribed genes, and its extensive colocalization with Tet1 is in agreement with their functional interaction. Moreover, MAB-seq analysis performed in Tdg-silenced cells revealed a significant increase of 5fC and 5caC levels on the TSSs of highly expressed genes.

In conclusion, we here presented a fast and reliable method to map 5fC and 5caC at a single-base resolution and provided a genome-wide map of these modifications in mouse ESCs. The presence of 5fC and 5caC residues on the promoters of highly expressed genes revealed the occurrence of an active DNA methylation and demethylation dynamics on transcriptionally active promoters, which is mediated by the interplay between the DNA methyltransferases Dnmt1 and Dnmt3a with the Tet1/Tet2 and Tdg proteins.

EXPERIMENTAL PROCEDURES

Cells Culture

Mouse embryonic stem cells E14 were cultured as previously described in Krepelova et al. (2014). Briefly, cells were grown on 0.1% gelatin-coated plates and maintained in DMEM (4.5g/l D-glucose), supplemented with 15% heat-inactivated fetal bovine serum, 0.1 mM nonessential amino acids, 1 mM sodium pyruvate, 0.1 mM 2-mercaptoethanol, 25 U/ml penicillin, 25 μg/ml streptomycin, and 1,500 U/ml leukemia inhibitory factor.

Antibodies

Antibodies were purchased from Imgenex (Dnmt1 IMG-261-A, Dnm3a IMG-268A, and Dnmt3b IMG-184A), Abcam (Tdg ab154192 and single-stranded DNA [ssDNA] AB78856), Millipore (Tet1 09-872), Diagenode (Tet2 mAb-179-050 and 5-hydroxymethyl cytosine mAb-31HMC-100), Active Motif (5-formyl cytosine V13-61223 and 5-carboxyl cytosine V13-61225), and Eurogentec (5-methyl cytosine BI-MECY-0500).

RNA Extraction and qRT-PCR

RNA samples were extracted directly from cultured cells using Trizol reagent (Gibco) followed by isopropanol precipitation. RNA was analyzed as previously described in Incarnato et al. (2013). Briefly, mRNA quantitation was performed by qRT-PCR using Superscript III platinum Onestep qRT-PCR System kit (Invitrogen) and normalized on Actb mRNA. Oligonucleotide sequences are indicated in Table S2.

MAB on DNA Standards

Five hundred nanograms of DNA standards were incubated with 1 U of M.SssI CpG methyltransferase (New England Biolabs), supplemented with 160 μM final S-Adenosyl methionine for 2 hr at 37°C, and then DNA was purified using QIAQuick PCR Purification Kit (QIAGEN) according to the manufacturer's instructions. Bisulfite conversion was performed using the EpiTect Bisulfite Kit (QIAGEN). DNA was amplified by PCR for ten cycles, cloned in TOPO-Blunt (Invitrogen), and verified by sequencing.

MAB-Seq

For 5fC and 5caC sequencing, 1 μg of mouse embryonic stem cell (mESC) genomic DNA (plus 1 ng of *E. coli* DH10B genomic DNA) was incubated with 4 U of M.SssI CpG methyltransferase (NEB), supplemented with 160 μM final S-Adenosyl methionine for 2 hr at 37°C. The incubation was repeated three times to maximize CpG methylation. For whole-genome MAB-seq, methylated DNA was sheared using a Bioruptor Twin (Diagenode) for two runs of 15 cycles (30 s on, 30 s off) at high power setting. For RRMAB-seq, methylated DNA was digested for 4 hr at 37°C with 200 U of MspI restriction endonuclease (NEB). Fragmented/digested DNA was then end-repaired, deoxyadenosine-tailed, and ligated to methylated adapters, using the Illumina TruSeq DNA Sample Prep Kit, following manufacturer instructions. Adaptor-ligated DNA was treated for additional three cycles with M.SssI enzyme to methylate the unmodified cytosines introduced by the end-repair step. DNA was loaded on an EGel Size select 2% agarose precast gel (Invitrogen), and a fraction corresponding to fragments ranging from 180 bp to 350 bp was recovered. Purified DNA was then subjected to bisulfite conversion using the EpiTect Bisulfite Kit (QIAGEN). Bisulfite-converted DNA was finally enriched by 15 cycles of PCR using Pfu Turbo Cx HotStart Taq (Agilent). All sequencing experiments were mapped on a recent genome reference assembly of E14 mESC to avoid false methylation calls due to the presence of single-nucleotide variants in the genome (Incarnato et al., 2014). Raw and analyzed data features are deposited in Table S1.

Dot Blot Analysis

For dot blot analysis, DNA was denatured with 0.4 M NaOH and then incubated for 10 min at 95°C before being spotted onto Hybond-N+ membranes (GE Healthcare). The membrane was UV-crosslinked for 10 min and saturated in 5% milk in PBS + 0.1% Tween 20 (Sigma-Aldrich) for 1 hr. ssDNA was used as loading control.

ACCESSION NUMBERS

Raw data for MAB-seq, RRMAB-seq, RRBS, and ChIP-seq were deposited to the NCBI GEO and are available under accession number GSE55660.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2015.01.008.

Figure 4. An Active Methylation/Demethylation Dynamics Occurs at Expressed Gene TSSs

(A) Heatmap representations on a window of ±5 kb around the TSS of genes for 5mC at different percentages. Plots of mean 5mC distribution in each range are represented under the respective heatmaps. Genes were ascending sorted according to levels of H3K4me3.

(B) Schematic overview of our experimental approach. ESCs are transfected with shRNAs targeting Tet1/Tet2, Dnmt1, Dnmt3a, and Dnmt3b. After knockdown, genomic DNA is extracted and subjected to reduced representation bisulfite sequencing (RRBS), which enables targeted analysis of DNA methylation on CpG-rich regions.

(C) Western blot of ESCs transfected either with a control shRNA (shCtr) or two shRNA-targeting Tet1 and Tet2. β-actin is used as loading control.

(D) Box plot of C methylation levels (on CpG detected as methylated by RRBS) on the promoter of highly expressed genes (upper quartile) in control versus Tet1/Tet2 double-knockdown ESCs. p value is given by Wilcoxon rank sum test statistics.

(E) Western blot of mESC transfected either with control shRNA (shCtr) or shRNAs targeting Dnmt1, Dnmt3a, and Dnmt3b. β-actin is used as loading control.

(F) Box plot of C methylation levels (on CpGs detected as methylated by RRBS) on the promoter of highly expressed genes (upper quartile) in control versus Dnmt1, Dnmt3a, and Dnmt3b single-knockdown and Dnmt3a/Dnmt3b double-knockdown ESCs. p values are given by Wilcoxon rank sum test statistics. n.s., not significant.

AUTHOR CONTRIBUTIONS

F.N., D.I., and S.O. conceived and designed the research. F.N., D.I., A.K., S.R., F.A. C.P., C.M., and F.D.B. performed experiments. F.N., D.I., and S.O. analyzed the data and wrote the paper.

ACKNOWLEDGMENTS

This work was supported by the Associazione Italiana Ricerca sul Cancro (AIRC) IG 2011 11982.

Received: November 10, 2014
Revised: December 24, 2014
Accepted: December 30, 2014
Published: February 5, 2015

REFERENCES

Bestor, T.H. (2000). The DNA methyltransferases of mammals. Hum. Mol. Genet. *9*, 2395–2402.

Booth, M.J., Branco, M.R., Ficz, G., Oxley, D., Krueger, F., Reik, W., and Balasubramanian, S. (2012). Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. Science *336*, 934–937.

Booth, M.J., Marsico, G., Bachman, M., Beraldi, D., and Balasubramanian, S. (2014). Quantitative sequencing of 5-formylcytosine in DNA at single-base resolution. Nat. Chem. *6*, 435–440.

Cortázar, D., Kunz, C., Selfridge, J., Lettieri, T., Saito, Y., MacDougall, E., Wirz, A., Schuermann, D., Jacobs, A.L., Siegrist, F., et al. (2011). Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. Nature *470*, 419–423.

Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D., et al. (2011). Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. Cell *146*, 67–79.

Costello, J.F., Frühwald, M.C., Smiraglia, D.J., Rush, L.J., Robertson, G.P., Gao, X., Wright, F.A., Feramisco, J.D., Peltomäki, P., Lang, J.C., et al. (2000). Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. Nat. Genet. *24*, 132–138.

Easwaran, H., Johnstone, S.E., Van Neste, L., Ohm, J., Mosbruger, T., Wang, Q., Aryee, M.J., Joyce, P., Ahuja, N., Weisenberger, D., et al. (2012). A DNA hypermethylation module for the stem/progenitor cell signature of cancer. Genome Res. *22*, 837–849.

Fatemi, M., Hermann, A., Gowher, H., and Jeltsch, A. (2002). Dnmt3a and Dnmt1 functionally cooperate during de novo methylation of DNA. Eur. J. Biochem. *269*, 4981–4984.

Feng, S., Jacobsen, S.E., and Reik, W. (2010). Epigenetic reprogramming in plant and animal development. Science *330*, 622–627.

Fouse, S.D., Shen, Y., Pellegrini, M., Cole, S., Meissner, A., Van Neste, L., Jaenisch, R., and Fan, G. (2008). Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. Cell Stem Cell *2*, 160–169.

Gowher, H., Stockdale, C.J., Goyal, R., Ferreira, H., Owen-Hughes, T., and Jeltsch, A. (2005). De novo methylation of nucleosomal DNA by the mammalian Dnmt1 and Dnmt3A DNA methyltransferases. Biochemistry *44*, 9899–9904.

Gu, H., Smith, Z.D., Bock, C., Boyle, P., Gnirke, A., and Meissner, A. (2011). Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. Nat. Protoc. *6*, 468–481.

Habibi, E., Brinkman, A.B., Arand, J., Kroeze, L.I., Kerstens, H.H.D., Matarese, F., Lepikhov, K., Gut, M., Brun-Heath, I., Hubner, N.C., et al. (2013). Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. Cell Stem Cell *13*, 360–369.

He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., et al. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science *333*, 1303–1307.

Incarnato, D., Neri, F., Diamanti, D., and Oliviero, S. (2013). MREdictor: a two-step dynamic interaction model that accounts for mRNA accessibility and Pumilio binding accurately predicts microRNA targets. Nucleic Acids Res. *41*, 8421–8433.

Incarnato, D., Krepelova, A., and Neri, F. (2014). High-throughput single nucleotide variant discovery in E14 mouse embryonic stem cells provides a new reference genome assembly. Genomics *104*, 121–127.

Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C., and Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature *466*, 1129–1133.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science *333*, 1300–1303.

Jair, K.-W., Bachman, K.E., Suzuki, H., Ting, A.H., Rhee, I., Yen, R.-W.C., Baylin, S.B., and Schuebel, K.E. (2006). De novo CpG island methylation in human cancer cells. Cancer Res. *66*, 682–692.

Jones, P.A. (2002). DNA methylation and cancer. Oncogene *21*, 5358–5360.

Kahramanoglou, C., Prieto, A.I., Khedkar, S., Haase, B., Gupta, A., Benes, V., Fraser, G.M., Luscombe, N.M., and Seshasayee, A.S.N. (2012). Genomics of DNA cytosine methylation in Escherichia coli reveals its role in stationary phase transcription. Nat. Commun. *3*, 886.

Kangaspeska, S., Stride, B., Métivier, R., Polycarpou-Schwarz, M., Ibberson, D., Carmouche, R.P., Benes, V., Gannon, F., and Reid, G. (2008). Transient cyclical methylation of promoter DNA. Nature *452*, 112–115.

Krepelova, A., Neri, F., Maldotti, M., Rapelli, S., and Oliviero, S. (2014). Myc and max genome-wide binding sites analysis links the Myc regulatory network with the polycomb and the core pluripotency networks in mouse embryonic stem cells. PLoS ONE *9*, e88933.

Li, E. (2002). Chromatin modification and epigenetic reprogramming in mammalian development. Nat. Rev. Genet. *3*, 662–673.

Lu, X., Song, C.-X., Szulwach, K., Wang, Z., Weidenbacher, P., Jin, P., and He, C. (2013). Chemical modification-assisted bisulfite sequencing (CAB-Seq) for 5-carboxylcytosine detection in DNA. J. Am. Chem. Soc. *135*, 9315–9317.

Maiti, A., and Drohat, A.C. (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. J. Biol. Chem. *286*, 35334–35338.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., et al. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766–770.

Métivier, R., Gallais, R., Tiffoche, C., Le Péron, C., Jurkowska, R.Z., Carmouche, R.P., Ibberson, D., Barath, P., Demay, F., Reid, G., et al. (2008). Cyclical DNA methylation of a transcriptionally active promoter. Nature *452*, 45–50.

Neri, F., Incarnato, D., Krepelova, A., Rapelli, S., Pagnani, A., Zecchina, R., Parlato, C., and Oliviero, S. (2013a). Genome-wide analysis identifies a functional association of Tet1 and Polycomb repressive complex 2 in mouse embryonic stem cells. Genome Biol. *14*, R91.

Neri, F., Krepelova, A., Incarnato, D., Maldotti, M., Parlato, C., Galvagni, F., Matarese, F., Stunnenberg, H.G., and Oliviero, S. (2013b). Dnmt3L antagonizes DNA methylation at bivalent promoters and favors DNA methylation at gene bodies in ESCs. Cell *155*, 121–134.

Neri, F., Dettori, D., Incarnato, D., Krepelova, A., Rapelli, S., Maldotti, M., Parlato, C., Paliogiannis, P., and Oliviero, S. (2014). TET1 is a tumour suppressor that inhibits colon cancer growth by derepressing inhibitors of the WNT pathway. Oncogene, Published online November 3, 2014. http://dx.doi.org/10.1038/onc.2014.356.

Raiber, E.-A., Beraldi, D., Ficz, G., Burgess, H.E., Branco, M.R., Murat, P., Oxley, D., Booth, M.J., Reik, W., and Balasubramanian, S. (2012). Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated

with transcription and depends on thymine DNA glycosylase. Genome Biol. *13*, R69.

Reik, W., and Walter, J. (2001). Genomic imprinting: parental influence on the genome. Nat. Rev. Genet. *2*, 21–32.

Shen, L., Wu, H., Diep, D., Yamaguchi, S., D'Alessio, A.C., Fung, H.-L., Zhang, K., and Zhang, Y. (2013). Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. Cell *153*, 692–706.

Song, C.-X., Szulwach, K.E., Dai, Q., Fu, Y., Mao, S.-Q., Lin, L., Street, C., Li, Y., Poidevin, M., Wu, H., et al. (2013). Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. Cell *153*, 678–691.

Teng, I.W., Hou, P.C., Lee, K.D., Chu, P.Y., Yeh, K.T., Jin, V.X., Tseng, M.J., Tsai, S.J., Chang, Y.S., Wu, C.S., et al. (2011). Targeted methylation of two tumor suppressor genes is sufficient to transform mesenchymal stem cells into cancer stem/initiating cells. Cancer Res. *71*, 4653–4663.

Williams, K., Christensen, J., Pedersen, M.T., Johansen, J.V., Cloos, P.A.C., Rappsilber, J., and Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature *473*, 343–348.

Wu, H., D'Alessio, A.C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Sun, Y.E., and Zhang, Y. (2011). Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. Nature *473*, 389–393.

Yu, M., Hon, G.C., Szulwach, K.E., Song, C.-X., Zhang, L., Kim, A., Li, X., Dai, Q., Shen, Y., Park, B., et al. (2012). Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. Cell *149*, 1368–1380.